

Chapitre I - Exposé de la problématique

Ce chapitre a pour objectif de présenter un cadre de recherche dans lequel la question des fondements théoriques de l'indexation puisse être posée.

En effet, le sujet de notre recherche n'est pas si évidemment pertinent : il suppose une approche de l'objet d'étude - l'indexation - et un type d'analyse - permettant de capter les fondements théoriques - qui coïncident peu avec la représentation de l'indexation telle qu'on la trouve traditionnellement exprimée dans la littérature.

Notre recherche repose en effet sur deux présupposés :

- d'une part, l'indexation peut faire l'objet d'une étude spécifique, c'est-à-dire constituer, en elle-même, un objet de recherche (I) ;
- d'autre part, une approche de l'indexation qui se veut théorique ne peut donner corps à une théorie* de l'indexation ; elle peut en revanche tenter d'établir les fondements théoriques de l'indexation (II).

I- Définir l'objet d'étude : approches de l'indexation

S'il paraît particulièrement essentiel de définir, dans cette recherche, un cadre pour penser notre objet d'étude, c'est que l'indexation en tant que telle ne peut constituer un objet de recherche dans la perspective de toutes les approches.

Après avoir situé la problématique de l'indexation comme objet d'étude, nous examinerons les propositions de la littérature classique ; nous en dégagerons les limites, à partir desquelles nous proposerons un autre mode d'appréhension de l'indexation.

I.1- L'indexation, un objet d'étude ?

Que l'indexation puisse constituer un objet d'étude ne va pas de soi ; la difficulté peut être ainsi exprimée :

- sur un plan méthodologique, un objet d'étude doit pouvoir être considéré de façon autonome : il faut pouvoir « isoler » un phénomène¹ ;
- sur un plan épistémologique, un objet étudié dans la perspective des sciences de l'information et de la communication est nécessairement un objet appréhendé dans sa finalité².

La question qui se pose est alors la suivante : quelle place accorder à la finalité de l'indexation ? Est-elle définitoire du processus de l'indexation, ou plutôt, comment est-elle définitoire de l'indexation ?

De ce point de vue, on peut opposer :

- une approche « instrumentale » : la définition de l'indexation est établie sur la base de la question « à quoi sert-elle ? ». Le point de vue sur l'objet est externe. Dans cette approche, l'indexation peut difficilement être pensée comme un objet d'étude dans la mesure où elle n'est pas « isolable » ;
- une approche « procédurale » : la définition de l'indexation est établie sous l'angle de la question « comment fonctionne-t-elle ? ». Le point de vue sur l'objet est interne. La finalité de l'indexation se laisse alors déduire de son mode de fonctionnement : c'est l'indexation elle-même qui porte son « mode d'emploi », la possibilité de son usage, sa finalité. À ce titre, elle peut constituer un objet d'étude spécifique. C'est l'hypothèse que nous défendrons.

I.2- Approches classiques de l'indexation

Les approches classiques* proposent une définition « instrumentale » de l'indexation qui peut prendre la forme suivante : l'indexation « a pour but de faciliter l'accès au contenu d'un document ou d'un ensemble de documents à partir d'un sujet ou d'une combinaison de sujets (ou de tout autre type d'entrée utile à la recherche) »³. Traditionnellement, l'indexation est définie de cette façon : comme

¹. Granger [1993, p. 72] parle de « réduction des phénomènes aux objets de science ».

². Rappelons que les sciences de l'information et de la communication s'attachent à « l'étude des processus de l'information et de la communication relevant d'actions organisées, finalisées », Comité National d'Évaluation 1993, p. 123 (c'est nous qui soulignons). La notion d'« action finalisée » peut être entendue comme « action née d'un besoin social », Granger 1993, p. 33.

³. Pomart et Sutter 1997, p. 284. Ce type de définition se trouve aussi dans le discours normatif (voir ci-après) mais aussi dans les discours didactiques [par exemple, Chaumier 1996, p. 18 : « Description du contenu du document à l'aide de mots clés (ou indices de classification) pour faciliter la mémorisation du contenu de ce document pour une recherche ultérieure »] et

un outil au service d'une fonction, la recherche documentaire*.

Le problème vient de ce que la finalité de l'indexation, lorsqu'elle est formulée par la notion de recherche documentaire, introduit une « circularité » d'analyse : la recherche documentaire est, en effet, elle-même définie par rapport à l'indexation.

La circularité de cette approche est particulièrement visible dans le texte de la norme :

- l'une des finalités de l'indexation est de permettre la recherche d'information :
« la finalité de l'indexation est de permettre une recherche efficace des informations contenues dans un fonds de documents et d'indiquer rapidement, sous forme concise, la teneur d'un document » ;
 - la recherche d'information est contrainte par l'indexation :
« l'indexation conduit à l'enregistrement des concepts contenus dans un document, sous une forme organisée et facilement accessible, c'est-à-dire à la confection d'outils de recherche documentaire. [...] La recherche des informations enregistrées [...] s'opérera à partir de ces outils de recherche documentaire ».
- Norme AFNOR Z 47-102 (1978), p. 225.

On comprend aisément pourquoi l'indexation est ainsi définie par les professionnels : l'indexation représente, pour eux, un moyen de remplir leur mission qui est, entre autres, de permettre de retrouver des documents.

L'indexation intéresse d'abord le professionnel sous l'angle de ce à quoi elle sert.

Cependant, il n'est pas sûr que cette approche « instrumentale » de l'indexation permette de l'étudier, ou du moins, de la constituer comme objet d'étude. En effet, l'approche de la finalité de l'indexation en termes de recherche documentaire s'avère problématique :

- le modèle de la recherche documentaire qui sert implicitement de référence à ce type de définitions se révèle être un modèle partiel (A) ;
- la recherche documentaire apparaît en outre comme une finalité « seconde » de l'indexation : c'est la notion plus large de « service à rendre » qui constitue le fond des approches classiques de l'indexation (B) ;
- la question de la recherche documentaire s'avère n'être enfin qu'une « inversion » de celle de l'indexation : la symétrie qui se dessine entre indexation et recherche documentaires conduit à rendre les objets indistincts (C).

dans les discours techniques [par exemple, Menon 1988, p. 146 : « Identification et enregistrement des unités d'information minimales pertinentes **pour apporter des réponses aux requêtes présentées au système d'information** »]. C'est nous qui soulignons.

A- Un modèle partiel de la recherche documentaire : le modèle de l'« *Information Retrieval* »

L'approche classique de l'indexation qui met en avant, comme l'indique la norme, la « confection des outils de recherche » (c'est-à-dire la confection de langages documentaires*), adopte implicitement un modèle de la recherche documentaire fondé sur la technique de l'appariement entre les mots d'une requête et les mots issus de l'analyse d'un document. Comme l'a notamment mis en valeur Kolmayer⁴, cette technique de l'appariement relève du modèle de l'*Information Retrieval*, modèle dominant dans les années 1970, qui voient se développer les systèmes d'interrogation informatisés. Or depuis, et notamment grâce aux analyses cognitives des situations de recherche d'information⁵, ce modèle de la recherche documentaire s'est révélé inadéquat : il est notamment apparu que le besoin d'information, ne restant pas constant au cours d'une session de recherche, ne pouvait se trouver exprimé une fois pour toutes dans une requête. C'est ainsi que, depuis le milieu des années 1980 surtout, le problème de la recherche documentaire a pu être reformulé :

S'agissant de l'accès à l'information, les dernières années ont vu la problématique de l'appariement d'une requête à un ensemble de documents se positionner progressivement à l'intérieur d'un cadre plus vaste, celui de la satisfaction du besoin d'information de l'utilisateur. La recherche d'information est devenue alors un problème de recherche cognitive.
Dachelet, 1990, p. 24.

Il peut paraître excessif (et un peu rapide) de considérer la recherche documentaire comme relevant exclusivement d'une approche cognitive ; cependant, il semble tout à fait nécessaire de prendre en compte les résultats des analyses cognitives des situations de recherche : si la pratique d'indexation ne se justifie que pour rendre possible la recherche documentaire, comment pourrait-elle ignorer ces nouvelles représentations des situations d'interrogation ? L'indexation telle qu'elle se pratique reste-t-elle compatible avec ce que l'on connaît désormais des modes de recherche ? Ne se trouve-t-elle pas remise en cause et amenée à se redéfinir ? C'est le sens de l'interrogation que formule Sylvie Lainé-Cruzel dans ces termes :

L'approche dominante concernant l'indexation postule que le sens est contenu dans le document et que, pour aboutir à une représentation du sens, il faut analyser aussi finement que possible la structure apparente du texte, c'est-à-dire identifier et caractériser (par des outils linguistiques et statistiques) la forme prise par le texte, comme étant le reflet exact du sens que l'auteur a donné au texte. Parallèlement émergent de nouveaux modèles, encore au stade théorique, qui s'appuient sur l'idée que le sens est construit par le lecteur, qu'il est différent d'un lecteur à l'autre, et

4. Kolmayer 1995, p. 27-33.

5. Pour la bibliographie, on peut se reporter à Kolmayer 1997.

qu'une bonne représentation du contenu passe par la prise en compte de certaines caractéristiques du futur utilisateur. Les approches sont-elles complémentaires ? antinomiques ?

Lainé-Cruzel, 1994, p. 143.

Si l'on peut admettre que la norme pour l'indexation des documents⁶, élaborée à des fins professionnelles en 1978, ait pu prendre implicitement pour cadre le modèle dominant de la recherche documentaire à l'époque (celui de l'*Information Retrieval*), on peut être plus surpris de relever que ce type de représentation - partiel - des situations de recherche documentaire n'ait pas conduit depuis à une redéfinition de l'indexation.

Cependant, à y regarder de près, il apparaît que la finalité de l'indexation exprimée en termes de recherche documentaire n'est, en fait, que « seconde ». C'est implicitement dans le cadre d'un objectif de communication plus large que l'indexation est en réalité définie. C'est cet objectif qui contraint le modèle de la recherche documentaire à être réduit, dans les approches classiques, à celui de l'*Information Retrieval*.

B- Hypothèse implicite sur les objectifs de la communication : l'« hypothèse de service »

La formule « hypothèse de service » a été proposée par Escarpit⁷ pour exprimer une des limites du modèle mécaniste de l'information.

En montrant que l'approche classique de l'indexation met en œuvre cette hypothèse de service, on pourra mettre au jour le modèle mécaniste dans lequel implicitement elle s'inscrit.

Rappelons que le modèle mécaniste de l'information, qui repose sur une formule mathématique que Shannon a développée en 1948, propose une représentation linéaire de la communication, dans laquelle le transfert de l'information s'effectue de la source au destinataire de la façon suivante :

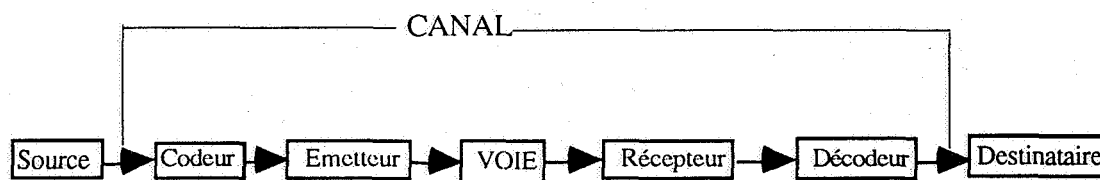


Figure 1 - Le schéma linéaire de la communication
Escarpit 1991, p. 27.

⁶. Norme AFNOR Z 47-102 (1978).

⁷. Escarpit 1991, p. 44-46.

Dans cette représentation, seul le destinataire est concerné par la signification des messages transmis. Celui qui véhicule les messages doit être, comme l'est un employé des Postes⁸, insensible au contenu : le messenger est simplement chargé de rendre un service, c'est-à-dire de « transmettre le plus rapidement, le plus économiquement et surtout le plus fidèlement possible au destinataire l'information contenue dans les messages de la source »⁹.

Cette notion de service à rendre devient une hypothèse - et une hypothèse douteuse, dit Escarpit¹⁰ - quand elle est sollicitée par les tenants du discours mécaniste pour ne pas avoir à s'exprimer ni sur la source de l'information (quelle est-elle ? elle-même de l'information ?), ni sur le message, ou plutôt sur la forme du message (quelle est sa nature ? symbolique ?). Autrement dit, et la critique d'Escarpit porte sur ce point, « ce que voudrait faire accepter ce que nous nommons l'hypothèse de service : [c'est que] l'employé des Postes qui achemine le télégramme n'est pas responsable de ce qui se passe dans l'esprit de l'expéditeur ou du destinataire »¹¹. L'hypothèse de service revient donc à poser que le transfert de l'information ne participe lui-même ni à la fabrication de l'information ni à l'interprétation que le destinataire peut en faire¹².

En examinant les définitions classiques de l'indexation, on relève la présence de l'hypothèse de service mise en lumière par Escarpit :

- l'indexation se donne comme une réponse technique à un problème technique de transmission d'information, sans qu'elle ait à se préoccuper du contenu de ce qu'elle transmet. L'enjeu de l'indexation a en effet pu être ainsi exprimé : « comment dans l'expression d'un sujet préserver la similitude du sens à travers la variété et les incertitudes des langues naturelles ? »¹³ ; or ni la question du « sujet » ni celle du « sens » ne sont véritablement posées ;
- l'approche classique ne s'exprime pas sur la source de la transmission (l'indexeur) pas plus que sur la forme du message (les langages documentaires) : ni l'indexeur ni les langages documentaires n'apparaissent comme des éléments susceptibles

⁸. Escarpit [*Ibid.*, p. 30-31] reprend ici l'image qu'a proposée Roubine dans son Introduction à la théorie de la communication [1970].

⁹. *Ibid.*, p. 45.

¹⁰. *Ibid.*, p. 44-45.

¹¹. *Ibid.*, p. 46.

¹². *Id.*

¹³. Maniez 1993, p. 254.

d'influer sur la réception de l'information. L'indexation se donne comme une opération neutre et transparente de transfert d'information.

Il apparaît donc que les approches classiques de l'indexation reposent implicitement sur une représentation particulière de la recherche documentaire dans la mesure où celle-ci permet de maintenir valide le modèle présupposé de la transmission transparente de l'information.

Or, si le modèle mécaniste a été reconnu essentiel pour rendre compte du fonctionnement du « canal » dans la chaîne de communication, il a été parallèlement jugé insuffisant pour considérer et la source et la forme du message¹⁴. Ces faiblesses ont conduit les chercheurs en sciences de l'information et de la communication à reprendre et à modifier ce modèle inspiré de Shannon ; cependant, toutes les approches de l'indexation ne semblent pas avoir intégré ces critiques. Est-il envisageable de considérer que cette non-prise en compte soit liée au caractère « vicié »¹⁵ de l'hypothèse de service qui, si elle met bien en avant la question du « service » à rendre, permet aussi de ne pas avoir à en dire plus ?

L'adoption implicite du modèle mécaniste s'avère en outre fâcheuse en ce qu'elle tend à rendre indistinctes indexation et recherche documentaires : les deux opérations se donnent en effet comme les deux faces d'un même processus¹⁶.

C- Hypothèse implicite de la symétrie

Les approches classiques de l'indexation admettent implicitement ce que nous appelons l'hypothèse de la symétrie, hypothèse qui permet de poser la notion d'« indexation des requêtes » et de la traiter au même titre et dans la même forme que l'indexation des documents¹⁷. Cette hypothèse conduit à rendre indistinctes indexation et recherche documentaires. Si l'indexation et la recherche constituent les deux faces d'un même processus, l'objet d'étude, dans le cadre des approches classiques, ne saurait être ni l'indexation ni la recherche documentaires mais le « processus » censé rendre compte de ces deux pratiques. Or, si ce processus existe, on ne le trouve ni nommé ni décrit dans la littérature du domaine¹⁸.

Il est clair que cette hypothèse de symétrie ne tient que dans le cadre des modèles

14. Escarpit 1991, p. 45.

15. Escarpit avance la notion de « duperie » pour qualifier les démarches qui établissent une « frontière entre le canal et ce qui vient avant et après lui ». *Ibid.*, p. 46.

16. Par exemple, Hodge [1992, p. 11] : « Indexing and information retrieval can be viewed as two sides of the same coin ».

17. Voir la norme AFNOR Z 47-102 (1978), p. 225 : l'« indexation de la question, opération analogue à l'indexation des fonds ».

18. Du moins dans la partie de la littérature du domaine que nous avons explorée, voir la bibliographie.

précédemment explicités :

- la technique de l'appariement impose en effet un traitement des questions dans les mêmes termes que ceux utilisés pour l'indexation des documents¹⁹ ;
- l'hypothèse de service dans laquelle s'inscrit l'indexation suppose une mise à distance du « contenu » véhiculé, mise à distance qui autorise un traitement indifférencié des documents et des questions, les deux étant appréhendés sans distinction en termes de message²⁰.

L'hypothèse de la symétrie, qui donne la possibilité de parler d'indexation de requêtes, apparaît donc plus comme une conséquence de l'adoption implicite du modèle mécaniste que comme une volonté positive de traiter sur un même plan requêtes et documents. Preuve en est que cette conséquence semble bien embarrassante aux yeux mêmes de ceux qui pratiquent l'indexation des requêtes.

On peut en effet relever, dans la littérature, un silence, évocateur nous semble-t-il, sur la question de l'indexation des requêtes. Si elle est toujours citée dans les manuels, traités théoriques et normes, c'est toujours secondairement, de façon succincte et par analogie avec l'indexation des documents²¹. En outre, les quelques auteurs qui la traitent explicitement ne manquent pas de relever l'absence de symétrie absolue entre les deux processus :

- ainsi Chaumier [1996], par exemple, met-il en avant, dans l'indexation des questions, l'importance de la « maïeutique », ceci pour répondre à la critique des professionnels se plaignant que « l'utilisateur ne sait pas ce qu'il veut » quand il formule une requête. L'indexation des questions nécessite, précise-t-il, un dialogue susceptible de permettre à l'utilisateur de trouver les « bons » mots ;
- pour les auteurs qui travaillent dans le cadre d'une automatisation de l'indexation, la symétrie de traitement est encore moins évidente.

¹⁹. AFNOR Z 47-102 (1978), p. 225 : « L'utilisation d'un langage documentaire pour ces deux opérations permet d'obtenir une coïncidence exacte du vocabulaire d'indexation des documents avec celui des questions auxquelles ces documents apportent une réponse ».

²⁰. Cette indifférence au contenu, propre au modèle mécaniste, est clairement exprimée par Roubine, cité in Escarpit 1991, p. 31 : « La signification des messages n'est pas prise en considération. [...] Il n'y a aucun inconvénient à identifier texte et message ».

²¹. Nous n'avons pas trouvé d'ouvrages ou d'articles traitant exclusivement l'indexation des requêtes. Par contre, dans les ouvrages ou articles abordant explicitement l'indexation des documents, on trouve toujours quelques phrases sur l'indexation des requêtes. L'analyse de la norme AFNOR est, de ce point de vue, éclairante : sous le titre « Principes pour l'indexation des documents », il est bien question, dans la quasi-totalité de la norme, de l'indexation des documents, à l'exception d'une phrase (p. 225) qui, presque incidemment, indique qu'un traitement analogue pourra être fait sur les questions des utilisateurs.

Ainsi Menon²², s'il assimile sous le terme de « texte » documents et requêtes, établit-il une différence de procédure : dans le cas du document, l'indexation porte sur des textes longs et doit être sélective ; dans le cas de la requête, l'indexation porte sur un texte court et doit être exhaustive.

L'hypothèse d'une symétrie, si elle est, au travers de ces deux exemples, quelque peu prise en défaut, n'est pas, on le voit, réellement mise en question, malgré les impasses où elle mène invariablement.

Les impasses apparaissent clairement dès lors qu'il devient nécessaire de distinguer les objets, ne serait-ce que pour les mesurer, les évaluer²³, notamment parce que l'on ne sait pas exactement qui, de l'indexeur ou de l'utilisateur d'un système d'information²⁴, est censé dégager le « contenu », les informations d'un document.

En effet, dans l'hypothèse de la symétrie de l'indexation, la construction de l'information n'est le fait de personne : elle est le fruit d'une rencontre, presque fortuite, entre deux ensembles de mots élaborés pourtant par des acteurs différents, animés d'intentions différentes. Malgré ces différences, l'indistinction entre documents et requêtes peut continuer à se penser, comme le note Danièle Dubois, grâce à « la notion d'information [qui] a ainsi la vertu d'unifier la sémantique des divers objets tant matériels que matérialisés ou mentaux »²⁵.

Quant à la rencontre et à la confrontation de ces différentes « informations » qui se donnent sous la même forme, quand elles ne sont pas purement et simplement ignorées, elles sont décrites comme relevant d'une « stratégie » qui ne concerne plus vraiment l'indexation :

L'essentiel dans un système documentaire est que la représentation des documents d'une part et la question d'autre part se rejoignent. Quant à définir à quel endroit se situe le point de jonction, cela est du ressort de la stratégie du système documentaire. Lallich-Boidin, 1986, p. 162.

En évacuant la question du « lieu de rencontre » entre documents et requêtes, l'hypothèse de la symétrie apparaît comme ce qui permet de ne pas poser la question de la construction de l'information²⁶.

²². Menon 1988, n. 7, p. 152.

²³. Voir, sur ce point, Sparck-Jones (ed.) 1981, p. 214 : « It is difficult to valid access indexing correctness without retrieval. Retrieval however is not a real solution : why should a particular set of queries be used to test indexing ? No indexers or judges can foretell all future uses of documents ».

²⁴. Dans la suite de ce document, on emploiera uniquement le terme « utilisateur » pour faire référence à l'utilisateur d'un système d'information (automatisé ou pas).

²⁵. Dubois 1995, p. 89.

²⁶. *Id.* : « La recherche documentaire et plus généralement les sciences de l'information reposent sur une théorie de l'information qui, à la différence d'une théorie sémantique, ne pose ni la question

L'hypothèse de la symétrie, à l'œuvre dans les approches classiques de l'indexation, permet donc d'évacuer, sans justification véritable, les questions de la place et du rôle des différents acteurs - indexeurs et utilisateurs - de l'indexation ; autrement dit, dans le cadre d'une approche classique, ces questions ne relèvent pas, semble-t-il, d'une étude de l'indexation.

Si l'on s'en tient au discours classique, il paraît difficile de conduire une étude de l'indexation en tant qu'objet autonome, hors du cadre de la recherche documentaire. On aurait néanmoins tort d'entériner une telle impossibilité sans préalablement examiner sur quoi elle repose.

L'analyse des présupposés du discours classique se révèle à ce titre éclairante. Elle permet de formuler les remarques suivantes :

- l'indexation est définie sous l'angle d'une finalité spécifique - la recherche documentaire - donnée comme extérieure à l'indexation. L'approche est en cela essentiellement « instrumentale » et tend à introduire une circularité d'analyse ;
- suffit-il de définir autrement la recherche documentaire pour dégager l'indexation d'une telle circularité ? La tentative serait, semble-t-il, insuffisante : l'approche classique adopte implicitement un modèle de communication (le modèle mécaniste) qui contraint la recherche documentaire à n'être qu'une procédure d'appariement ;
- faut-il alors considérer que l'indexation et la recherche fonctionnent comme les deux faces d'un même processus qui, lui, pourrait constituer un objet de recherche ? Force est de constater que cet objet semble bien impalpable, mais surtout qu'il conduit à entériner l'hypothèse que la fabrication de l'information ne concerne pas en propre le traitement documentaire : voilà une conclusion qui remettrait en cause le projet même des sciences de l'information et de la communication.

Parallèlement, l'examen des présupposés de l'approche classique de l'indexation permet de pointer les difficultés que rencontre une telle approche pour décrire précisément ce par quoi elle définit l'indexation : sa finalité. Le jeu de renvoi entre indexation et recherche, l'hypothèse opacifiante du « service à rendre », l'absence de prise en compte du rôle des différents acteurs, aboutissent chacun à repousser la

de la production du sens dans un texte, par un auteur ou un utilisateur formulant sa requête, ni celle, symétrique, de l'interprétation ».

question de la formulation précise et explicite de la finalité de l'indexation.
En quoi une approche non « instrumentale » de l'indexation peut-elle permettre de mieux l'appréhender ?

I.3- Pour une approche non « instrumentale » de l'indexation

La difficulté de constituer l'indexation comme objet d'étude autonome apparaît principalement dans le cadre de définitions qui lient intimement indexation et recherche documentaires.

Or on peut montrer que ces définitions véhiculent implicitement des hypothèses relevant de modèles jugés désormais partiels ou insuffisants. Ces modèles présentent en outre le fâcheux inconvénient de ne pas offrir de réponses claires aux questions suivantes : quelle est la finalité de l'indexation ? Quelle est la « stratégie » à mettre en œuvre pour qu'un utilisateur ait accès à des documents ? Il apparaît donc que l'on doit se dégager de ces définitions « instrumentales » ; mais apparaissent alors de nouvelles difficultés :

- d'une part, un paradoxe : peut-on étudier l'indexation en tant que pratique professionnelle finalisée tout en ne considérant pas cette finalité comme constitutive de la pratique ?
- d'autre part, une question : quelle est la finalité de l'indexation, si elle ne peut s'exprimer ni en termes de recherche documentaire ni en termes de « service » ?

Si l'on peut montrer que la finalité de l'indexation n'est pas à ce point ni spécifique ni « extérieure » à son mode de fonctionnement, alors une analyse de l'indexation comme objet d'étude autonome est possible.

Notre recherche conduit précisément à argumenter ces deux points : on pourra d'une part montrer que l'indexation peut se comprendre dans le champ plus vaste de la « diffusion des connaissances », où elle croise ce faisant d'autres pratiques professionnelles, comme celle de la vulgarisation scientifique²⁷. Nous proposerons un rapprochement de ces deux pratiques du point de vue de leur finalité. D'autre part, la finalité de l'indexation se laisse déduire de l'usage de la langue qu'elle met en œuvre : sur ce point, l'indexation rejoint la problématique classique de la mise en

²⁷. Voir, sur ce point, Jacobi [1987] qui propose de considérer les moyens de diffusion de la connaissance sous la forme d'un continuum : « La vulgarisation scientifique s'inscrit **dans un continuum de la diffusion de la science**, elle en est une des modalités. [...] La diffusion large, auprès d'un public indifférencié, par le moyen de rhétorique particulière n'est probablement qu'un cliché dénué de consistance. Dans les faits, c'est **une large panoplie de pratiques de socio-diffusion de la science** qu'il faudrait évoquer. » Jacobi 1987, respectivement p. 8 et p. 163. C'est nous qui soulignons.

relation entre les « mots » et les « choses », question à laquelle l'indexation propose une réponse « professionnelle », une réponse qui met en œuvre des moyens spécifiques à son champ d'exercice.

On fait donc l'hypothèse que l'on peut constituer l'indexation comme un objet d'étude autonome tout en rendant compte de la façon dont elle réalise son objectif : la finalité de l'indexation est alors inscrite dans le mode de fonctionnement même de l'indexation.

Cependant, à ce stade de la recherche, la possibilité d'une étude autonome de l'indexation ne peut être que présupposée. L'adoption de ce présupposé n'est pas sans conséquence.

Uniquement possible dans le cadre d'une approche non « instrumentale », la constitution de l'indexation comme objet d'étude autonome amène à modifier certaines des caractéristiques classiquement définitoires de l'indexation. Ces modifications portent essentiellement sur trois aspects :

- les objets de l'indexation : hors du cadre de l'hypothèse de la symétrie, l'indexation ne peut porter que sur les documents ; il ne sera pas question, dans cette recherche, de l'indexation des requêtes ;
- le processus de l'indexation : hors du cadre de l'hypothèse de service, l'indexation ne peut être comprise comme un simple transfert d'information ; elle sera plutôt appréhendée comme une opération, qui, à ce titre, produit et construit ses propres éléments (les documents et l'information, notamment) ;
- les outils de l'indexation : hors du modèle réductionniste de l'appariement, la problématique de l'indexation peut être dégagée de celle des langages documentaires ; on peut envisager d'autres moyens linguistiques par lesquels pourrait se réaliser l'indexation.

L'adoption d'une approche non « instrumentale » de l'indexation amène donc à évacuer, du champ de la recherche, l'indexation des requêtes, à mettre à distance le rôle des langages documentaires et à considérer au sens fort la notion de « processus » en indexation.

Le premier paragraphe de ce chapitre a permis de mettre en valeur le caractère problématique du statut de l'indexation comme objet d'étude autonome. Notre recherche s'appuie sur le présupposé qu'un tel statut peut néanmoins être établi.

II- Définir la méthode d'analyse : approches théoriques de l'indexation

Compte tenu de l'absence de spécificité accordée à la seule indexation dans les discours classiques, on ne s'étonnera pas d'y trouver peu d'approches théoriques spécifiques. N'est-on pas alors amené à interroger la notion même d'approche théorique de l'indexation ? Peut-on distinguer, dans ce cadre, « théorie » et « fondement théorique » (II.1) ?

Dès lors que l'indexation ne semble pouvoir constituer l'objet d'une théorie mais qu'elle trouve ses fondements dans plusieurs théories, quelle approche théorique peut-on légitimement privilégier ? En quoi une approche linguistique des fondements théoriques de l'indexation peut-elle constituer un point de vue pertinent (II.2) ?

Enfin, si l'indexation peut se concevoir comme un objet d'étude autonome, peut-elle constituer, en soi, un objet d'étude pour la linguistique ? Quelle méthode d'analyse permet d'étudier l'indexation sous l'angle d'une problématique linguistique (II.3) ?

II.1- Théories ou fondements théoriques de l'indexation ?

La possibilité de constituer une théorie de l'indexation est loin d'être partagée par tous les chercheurs du domaine. L'examen des productions scientifiques dessine principalement deux tendances :

- pour les uns, une théorie de l'indexation n'est purement et simplement pas envisageable²⁸ ;
- pour les autres, une théorie de l'indexation est envisageable, mais elle ne peut, en l'état actuel des connaissances, être élaborée²⁹.

On ne peut, à l'instar de Dachelet, que constater la grande absence des travaux théoriques sur l'indexation, absence qui contraste avec l'intense productivité des études concernant la recherche d'information :

Le travail qui s'accomplit en ce domaine [celui de la représentation du contenu des documents] est un travail de fond dont les résultats progressent lentement. Notre

28. C'est, par exemple, la position de Varet 1995, p. 630 : « L'indexation est une question complexe parce qu'elle ne comporte pas de solution universelle. [...] Il n'existe pas même de théorie générale permettant d'en maîtriser intellectuellement le schéma ».

29. Voir, par exemple, la position de Hutchins : « As our knowledge about the process of document representation remains inadequate, methods from AI and cognitive science are indeed called for : the process of document representation becomes an engineering problem ». Hutchins 1977 cité in Endres-Niggemeyer 1989, p. 231.

second thème : la recherche d'information, semble, par contraste, être l'objet d'une activité beaucoup plus intense et quasiment monopoliser les efforts de la recherche. Dachelet, 1990, p. 1.

Comment interpréter cette faiblesse de la recherche ?

En examinant les arguments avancés par ceux qui déniaient la possibilité d'une théorisation de l'indexation, on remarque que le principal argument mis en avant concerne l'utilité ou encore la pertinence d'une telle recherche³⁰, alors que, par contraste, les travaux sur la recherche documentaire semblent, eux, plus évidemment opératoires³¹.

Il est clair, et on ne saurait s'en étonner, que ces arguments et ce choix d'axes de recherche s'inspirent directement et exclusivement d'une approche purement « instrumentale » de l'indexation, propre au modèle mécaniste de l'information dégagé précédemment.

Mais ce modèle, que l'on peut qualifier de « vicié » en ce qu'il est repose sur l'hypothèse opportuniste du « service », conduit en fait à une contradiction :

- d'un côté, l'indexation n'y est vue que comme une technique, et à ce titre, elle n'intéresse les chercheurs que par les résultats qu'elle fournit ; cependant, en tant que technique, l'indexation devrait pouvoir faire l'objet d'une théorie. Or c'est une théorie de la recherche documentaire qui est établie ;
- d'un autre côté, l'indexation, jugée trop complexe, est implicitement assimilée à une pratique et se pose à ce titre comme non théorisable.

Pour rendre plus claire la confusion de statut (pratique/technique) qui, à nos yeux, caractérise l'indexation dans les approches qui la tiennent pour non théorisable, il importe de préciser ce que l'on peut entendre par « technique » et par « pratique »³².

³⁰. Voir, par exemple, la position de Lancaster 1991, p. 28 : « A number of "theories" of indexing have been put forward [...] but these tend not to be true theories and they offer little practical help for indexer ».

³¹. Voir, par exemple, cette remarque de Endres-Niggemeyer [1989, p. 230] : « This is indeed a very common attitude among information scientists that we do not need to know how indexers arrive at a particular description of the contents of a document ; all that matter is whether it enables users to find the document when required ».

³². Cette distinction nous est apparue essentielle à la lecture de Buisson 1995, notamment la problématique qu'elle pose p. 122 : « Doit-on parler d'une pratique de l'indexation ou d'une technique ? L'emploi de l'un ou l'autre terme n'est pas sans incidence sur la signification que l'on attribue à l'opération et sur le rôle que l'on assigne au documentaliste qui indexe ».

À la suite de Corinne Buisson³³, on reprend la distinction qu'Habermas³⁴ établit entre une technique et une pratique :

- une technique se comprend comme la version appliquée d'une théorie, elle est plus précisément « une application linéaire et univoque »³⁵ des lois ou des procédés que dégage une théorie ; c'est pourquoi ses résultats peuvent être tenus pour prédictibles et reproductibles à l'identique ;
- une pratique se réfère, elle, non pas à une théorie mais à un usage, à une norme ; elle tire son domaine de validité d'un système de valeurs, d'une « idéologie » entendue au sens de Boudon³⁶ : ses productions sont évaluées à cette aune.

Ces deux notions succinctement posées, on peut reformuler l'impasse et la contradiction auxquelles mène le modèle mécaniste de l'information : il incite à concevoir l'indexation comme une technique mais ne donne pas les moyens de restituer la théorie dont elle est censée être l'application ; comme la théorisation de l'indexation s'avère impossible, on pose alors implicitement que l'indexation est une pratique, et donc qu'elle ne peut faire l'objet d'aucune théorie.

La mise en valeur de cette confusion de statut peut suggérer une réponse à notre interrogation sur la rareté, voire l'inexistence, de théories de l'indexation, alors que pourtant certains la considèrent possible : l'indexation relèverait moins d'une technique, supposant une théorie, que d'une pratique, reposant sur un système de valeurs, une idéologie. Dès lors, il faut renoncer à toute tentative de théorisation et considérer pleinement l'indexation sous l'angle d'une pratique³⁷. Là encore, seules des approches qui s'écartent du modèle mécaniste peuvent tenir l'hypothèse de l'indexation comme pratique ; notre adoption d'une approche non « instrumentale » nous permet de la penser comme telle.

Notre recherche posera donc comme hypothèse que l'indexation constitue une pratique.

³³. Buisson 1995, p. 121-128.

³⁴. Habermas 1973 [1968].

³⁵. Habermas 1973 [1968], p. XXXV.

³⁶. Boudon 1986, p. 82 : l'idéologie peut être vue comme une « interprétation significative du monde échappant au critère de la vérité et de l'erreur et pouvant être expliquée à partir de l'environnement social de l'acteur ».

³⁷. Comme le note Corinne Buisson [1995, p. 121-123], il est rare que l'indexation soit posée, jusque dans ses ultimes implications, comme une pratique : la notion de « techniques documentaires » reste dominante.

L'indexation considérée comme une pratique impose un type de relation particulier aux champs théoriques. Cette relation se joue sur le mode de la **représentation*** : une pratique professionnelle comme celle de l'indexation repose, de façon implicite, sur une certaine représentation d'une ou plutôt de plusieurs théories (une théorie du langage, une théorie de la cognition, une théorie de la communication, etc.). C'est en ce sens que, selon nous, une étude scientifique de l'indexation ne peut être qu'une étude de ses fondements théoriques, c'est-à-dire qu'une étude de la représentation des théories sur lesquelles implicitement elle se fonde. Mais une étude nécessairement critique, dans la mesure où toutes les représentations théoriques sous-jacentes à l'indexation ne peuvent être comprises comme des fondements théoriques, c'est-à-dire comme des fondements valides du point de vue d'une théorie constituée.

L'étude des fondements théoriques de l'indexation se dédouble donc nécessairement en deux volets :

- une analyse descriptive des représentations théoriques sous-jacentes à la pratique d'indexation ;
- une analyse critique évaluant les distorsions (éventuelles) qui s'observent entre théorie scientifique et théorie représentée en indexation.

Pour mener cette étude des fondements théoriques de l'indexation ainsi entendue, devra être élaborée une méthodologie d'étude qui permette de confronter les deux niveaux en présence : le niveau relatif à la théorie sollicitée en indexation et le niveau relatif à la représentation qui est faite de cette théorie dans le cadre et pour les besoins d'une pratique.

Avant d'aborder la question d'une telle méthodologie, nous devons déterminer le type de théorie à partir de laquelle la pratique d'indexation construit ses représentations. À la suite d'autres auteurs³⁸, nous faisons l'hypothèse que les fondements théoriques de l'indexation, comme l'ensemble des objets d'étude des sciences de l'information et de la communication, relèvent de plusieurs champs disciplinaires³⁹. L'un de ces champs est celui des sciences du langage* : est-ce le plus essentiel, le plus déterminant ? Nous l'ignorons⁴⁰ ; mais il importe de noter que le champ de la linguistique est celui qui apparaît le plus visiblement sollicité dans la pratique de l'indexation (voir ci-après, § II.2).

³⁸. Miège par exemple 1993-1994.

³⁹. Voir, sur ce point, Le Coadic 1994.

⁴⁰. Blair [1990, p. 122] fait l'hypothèse que toute approche de l'indexation repose sur un modèle du langage : « Any theory of indexing or document representation presupposes a theory of language and meaning ».

Qu'est-ce que les sciences du langage nous permettent de voir de l'indexation et qu'est-ce qu'elles ne nous permettent pas d'atteindre ? Si seule une approche interdisciplinaire pourrait permettre de le dire précisément, nous essaierons néanmoins de circonscrire les objets de l'indexation dont une approche linguistique peut rendre compte (§ II.3).

II.2- Pour une approche linguistique des fondements théoriques de l'indexation

Si elle peut paraître relativement « naturelle »⁴¹, une approche linguistique de l'indexation ne présente cependant aucun caractère d'évidence. En effet, ce n'est pas parce que l'indexation manipule, entre autres, des objets de nature linguistique (textes des documents, mots des langages documentaires par exemple) qu'elle manipule des objets de la linguistique. En effet, sans faire une épistémologie des sciences du langage, on doit relever que la linguistique, en tant que discipline scientifique, d'une part n'est pas une (il y a plusieurs types de linguistique⁴²) et, d'autre part, qu'elle n'a pas pour ambition de rendre compte de tout le langage ni de toutes ses réalisations⁴³. De façon générale, on peut dire que l'objet de la linguistique⁴⁴ est la description de ce qui constitue une langue, c'est-à-dire, entre autres, la mise au jour de propriétés qui permettent de distinguer une langue d'une non-langue. Dans ce cadre, les objets que manipulent généralement les linguistiques sont des objets qui ne peuvent, à strictement parler, être traités que par elles et qui n'intéressent qu'elles⁴⁵.

Cependant, du point de vue qui est le nôtre (celui de l'analyse des représentations de la langue dans la pratique d'indexation), la linguistique, entendue comme « un ensemble de postulats sur la langue (ou le langage) et de méthodologies de description »⁴⁶, nous paraît pertinente, et ce à deux niveaux :

- au niveau des descriptions des faits de langue : les propriétés que la linguistique dégage à partir de l'étude de ses propres objets peuvent, moyennant une décontextualisation et une généralisation, expliquer

41. Par exemple Menon, 1988, p. 165 : « Les textes, en tant qu'objets langagiers, posent avant tout - et qui s'en étonnera - des problèmes linguistiques ».

42. Milner 1978, p. 11.

43. Milner 1989, p. 38-50.

44. Voir Milner *Ibid.*, p. 43 et suiv.

45. Milner fait observer (*ibid.*, p. 34) que « dès qu'elle dépasse la banalité, une proposition linguistique concerne peu de données à la fois et elle fait apparaître généralement ce que l'opinion courante tiendrait pour des détails ».

46. Marandin 1979, p. 18.

le fonctionnement d'un certain nombre de mécanismes à l'œuvre dans l'indexation, mécanismes peu visibles hors du prisme de l'analyse linguistique ;

- au niveau des postulats sur la langue* et le langage* : la théorie linguistique, si elle n'est pas la seule à tenir des propositions générales sur la langue et le langage, fournit l'un des cadres possibles pour discuter la conception du langage sous-tendue par la pratique d'indexation ; à ce titre, on pourra confronter la conception du langage telle que la véhicule l'indexation à la conception du langage postulée par les modèles linguistiques.

On peut, à titre d'exemple, citer quelques caractéristiques qui, issues de la pratique d'indexation, peuvent être appréhendées dans le cadre d'une approche linguistique, à l'un ou à l'autre des deux niveaux dégagés.

(i) Au niveau des descriptions des faits de langue

On liste ci-dessous, sans souci d'exhaustivité ni de détails, quelques particularités de l'indexation ou des langages documentaires qui recourent des descriptions de faits de langue menées en linguistique.

Convention d'écriture ou propriété des unités lexicales ?

Les langages documentaires utilisés en indexation sont constitués exclusivement de noms⁴⁷. Dans les termes de l'approche classique, il s'agit simplement de se donner là une convention d'écriture⁴⁸. Or, sur un plan linguistique, les unités lexicales de catégorie nominale sont des unités linguistiques qui présentent la propriété de pouvoir, en discours, référer à un objet. Nous reviendrons largement sur cette propriété linguistique qui fait apparaître la notion cruciale en indexation de référence*, cette notion ne pouvant apparaître en tant que telle si l'on se place du seul point de vue de la pratique.

Synonymie linguistique ou synonymie documentaire ?

La notion de synonymie qui relève du champ de la linguistique est utilisée en indexation dans une acception qui se veut particulière : on parle en effet de

47. Au sens de l'anglais *noun*, catégorie grammaticale, qui s'oppose à *name*, le « nom » du référent : le terme français ne rend pas cette distinction.

48. Voir, par exemple, la norme AFNOR Z 47-100 (1981), p. 187 : « La forme grammaticale du descripteur doit se conformer aux règles suivantes : **forme substantive** [...] », ou encore Chaumier 1978, p. 31 : « Les descripteurs sont des termes normalisés, les **règles d'écriture** d'un descripteur étant les suivantes : **forme substantive** ». C'est nous qui soulignons.

synonymie documentaire, que l'on oppose à synonymie linguistique, lorsque l'on veut décrire la possibilité de rapprocher des termes alors « que leur signification sémantique [est] différente »⁴⁹. Une description linguistique permet justement de montrer que la synonymie ne joue pas sur le sens des unités mais sur leur référent, ouvrant ici une nouvelle problématique pour l'étude de la référence en indexation : celle du rapport entre le sens et la référence, nous y reviendrons. Là encore cet enjeu ne peut apparaître clairement si l'on reste au niveau du seul discours classique.

Les « formes nouvelles de composition nominale »⁵⁰

Dans son article de 1966, Benveniste relevait l'« extension considérable » d'un type de termes « construit[s] sur un modèle qui n'est plus celui de la composition classique » : il s'agit de ce qu'il nomme les synapsies, qui sont du type « modulation de fréquence », « avion à réaction »⁵¹. Benveniste prédit que ce type de termes « sera la formation de base dans les nomenclatures techniques »⁵². Depuis, des études ont été menées en linguistique qui portent sur la description de telles unités : nous évoquerons l'une d'entre elles. On peut relever qu'à la même époque (dans les années 1970) apparaissent, dans les thésaurus, de nouveaux types de descripteur ayant précisément pour particularité d'être constitués de plusieurs « mots » (de type « contrôleur de gestion », « énergie géothermique »⁵³). Dans le discours classique, l'introduction de telles expressions ne se justifie que par rapport aux difficultés que rencontre une interrogation par unitermes⁵⁴. Par le biais de la description linguistique de la synapsie, on pourra montrer que cette apparition d'une nouvelle forme de descripteur peut relever d'une tout autre raison que celle de la seule performance en matière de recherche documentaire : une unité telle que la synapsie révèle en effet des propriétés qui intéressent au premier chef l'indexation elle-même et non simplement la recherche documentaire.

⁴⁹. Chaumier 1978, p. 33.

⁵⁰. Titre d'un article de Benveniste paru en 1966 et repris dans le chapitre XII de Benveniste 1974, p. 163-176.

⁵¹. Exemples repris de Benveniste. *Ibid.*, p. 172.

⁵². *Id.*

⁵³. Exemples repris de Chaumier 1978, p. 23.

⁵⁴. Chaumier, *id.* : « Devant les problèmes posés par l'utilisation de la post-coordination totale sur les unitermes au moment de l'indexation, un certain degré de précoordination se fit jour dans les thésaurus avec l'emploi des descripteurs. Il s'agit ici de précoordination au niveau du concept. C'est ainsi que les thésaurus admirent des expressions précoordonnées telles que “ contrôleur de gestion », “ énergie géothermique ” composées chacune de deux unitermes ».

(ii) *Au niveau des postulats sur la langue et le langage*

Si l'on ne prétend pas que seules les théories linguistiques sont à même de discuter les postulats que l'on peut tenir sur la langue et sur le langage, on tient que la conception du langage qui les constitue comme science peut fournir un référentiel pertinent pour penser les représentations du langage qui sous-tendent les pratiques d'indexation, et notamment :

La représentation de la langue comme une nomenclature

C'est l'adoption de cette représentation particulière du langage qui, semble-t-il, autorise la pratique documentaire à créer un langage artificiel à partir de la langue « naturelle ». Sur la base d'une critique de la vision de la langue comme nomenclature, la linguistique invite à repenser cette question et à substituer à la notion de création de langages documentaires celle de création d'une utilisation documentaire de la langue « naturelle »⁵⁵.

La représentation de la relation entre les mots et les choses

Cette relation entre les mots et les choses, dite aussi relation de référence, est au cœur des problématiques de la linguistique. L'une des façons de la problématiser peut être : « comment le langage parvient-il à parler du réel ? »⁵⁶. Par contraste, la relation référentielle n'est qu'implicitement présente dans le discours classique : cependant, nous pourrions observer que la problématique de la référence, une fois posée, apparaît en fait dans l'indexation sous des formes variées et qu'ainsi mise en valeur, elle permet de clarifier une partie du fonctionnement et de la finalité de l'indexation.

La question du sens

Là encore, et il n'est pas superflu de le noter, la question du sens n'est pas explicitement posée en indexation⁵⁷. En revanche, le discours classique manipule largement la notion de « contenu » (d'un document, d'une requête). Cette terminologie n'est pas neutre, comme nous tenterons de le montrer : elle suppose

⁵⁵. Milner 1989, p. 35 : « En fait, le problème doit être posé autrement : en tant que la science les saisit, les langues et le langage ne sont pas des matières réalisées ; ce sont plutôt les *lois* qui régissent ces " matières ". Inversement, les techniques de la langue n'ont pas pour fin de produire de nouvelles entités de langue (de nouveaux mots, de nouvelles structures, de nouvelles langues, etc.), mais de nouveaux objets où les langues *telles qu'elles sont* interviennent : ce ne sont donc pas à proprement parler les langues qui sont visées, mais les réalisations de langue - textes, messages, slogans, discours, etc. ».

⁵⁶. On reprend ici une formulation proposée dans Kleiber 1981, p. 11.

⁵⁷. Dubois 1995, *supra* (note 26).

une certaine conception du sens*, que les approches linguistiques, supposant souvent en des termes différents qu'il y a de la signification* dans et par la langue, nous permettront de mettre au jour.

Si ces quelques exemples peuvent montrer que la référence à la linguistique en tant que science est bien à l'œuvre dans la pratique de l'indexation, du moins au niveau implicite des représentations, par quelle méthode mener une étude qui permette à la fois de :

- montrer que ces représentations linguistiques peuvent être, pour certaines d'entre elles, considérées comme des fondements théoriques de l'indexation,
- maintenir distincts les deux types d'objet que nous aurons à manipuler - les objets de la linguistique et les objets de la pratique documentaire ?

La principale difficulté tient à ce que, la linguistique n'ayant rien de particulier à dire sur l'indexation, nous serons nécessairement amenée à procéder à des déplacements (sous forme de décontextualisations et/ou des généralisations) qui introduiront nécessairement des décalages de niveaux : le niveau de l'analyse linguistique n'est pas celui qui est directement pertinent pour l'indexation. Ainsi, traiterons-nous, par exemple, de la référence à partir d'études linguistiques ayant pour objet l'anaphore* ; de même aborderons-nous la question de la signification lexicale par le biais de l'étude morphologique d'un type de mots particuliers, les mots construits*, etc.

En outre, et ce second point constitue une autre limite importante de cette recherche, nous ne pourrions nous permettre de discuter la pertinence des descriptions linguistiques sollicitées, pouvant ainsi donner l'impression de ne retenir des linguistiques que la description de faits qui concerne de façon opportuniste notre objet.

On peut essayer de réduire l'incidence de ces deux limites en disposant d'une méthode d'analyse qui permette de donner un statut à ce type d'analyse du décalage entre modèle de langue et modèle d'usage de la langue. On attend de la méthode décrite ci-dessous qu'elle constitue un garde-fou évitant de céder à la tentation de plaquer une description linguistique sur des objets documentaires. On attend également de cette méthode qu'elle puisse accueillir d'autres types d'approches linguistiques que celles adoptées dans cette recherche : on espère disposer là d'un cadre qui permette aux modèles linguistiques que nous retiendrons de pouvoir être discutés dans le cadre particulier des problématiques de l'indexation.

II.3- Une méthode d'analyse du décalage

Nous présentons ci-après, en nous appuyant sur Berrendonner et Reichler-Béguelin [1989], une méthode d'analyse valable pour tout type de démarche qui cherche à confronter les propositions d'un modèle théorique aux réalisations d'une pratique qui paraît s'inspirer de ses objets et/ou de ses présupposés (A).

Nous préciserons ensuite la forme que prend cette méthode appliquée à notre approche linguistique de l'indexation. Dans le cadre de cette méthode, nous identifierons les objectifs que notre recherche peut se donner et les objets sur lesquels elle peut porter (B).

Nous aborderons pour finir à la fois les contours et les limites que suppose notre recherche (C).

A- Présentation de la méthode d'analyse du décalage

La perspective dans laquelle Berrendonner et Reichler-Béguelin [1989] étudient les pratiques de segmentation d'un texte (en lettres, mots, phrases) s'inscrit dans un cadre méthodologique plus global qui distingue les « représentations formelles » d'une part et les « catégorisations pratiques » d'autre part.

Les représentations formelles sont celles que produit la science, les catégorisations pratiques sont celles que produisent les pratiques : d'un côté, des représentations établies selon des critères scientifiques ; de l'autre, des catégorisations régies par des contraintes utilitaires. Les deux types d'entité se distinguent sur plus d'un point : homogénéité *versus* hétérogénéité des propriétés, objet à un *versus* plusieurs critère(s), indéformabilité *versus* adaptabilité des classes d'objets. C'est pourquoi il est courant de noter que « toutes les catégorisations pratiques diffèrent des représentations produites par la science : almanach vs météorologie, classification sociale des aliments vs diététique »⁵⁸.

Peut-on aller plus loin que la seule notification de cette différence ? Et, si oui, comment procéder ? Quelle méthode élaborer pour étudier les décalages entre représentations formelles et catégorisations pratiques, sans les exclure ni les dissoudre les unes aux autres ? Berrendonner et Reichler-Béguelin adoptent une méthode qui repose sur la confrontation non des entités elles-mêmes mais des **modèles** dans lesquels elles s'intègrent ; ces modèles ne sont pas tout à fait de même nature.

⁵⁸. Berrendonner et Reichler-Béguelin 1989, p. 106.

En effet, si la constitution d'un modèle répond à une exigence scientifique dans le cas des représentations formelles, elle répond à une nécessité d'usage dans le cas des catégorisations pratiques. La nature de chacun des deux modèles est donc foncièrement différente :

En matière d'interactions sociales, il importe que les divers usagers d'un même système disposent d'emblée d'un minimum de **représentations communes**, faute de quoi le temps d'agir se perd à négocier un accord des partenaires sur la « façon de voir les choses ». Ce besoin de fonder la coopération sur une base de représentations communes explique que les catégorisations pratiques empruntent volontiers à une **doxa**, à une tradition collective, ou à un fonds idéologique majoritaire des schèmes cognitifs stéréotypés, qui bénéficient de l'immédiateté de l'évidence. Ceux-ci sont alors ressentis, et généralement dépeints, comme un corps de normes sociales imposées aux individus.

Berrendonner et Reichler-Béguelin, 1989, p. 109. C'est nous qui soulignons.

On appellera la *doxa* d'une pratique « modèle d'utilisation », que l'on opposera au modèle scientifique, nommé alors « modèle de fonctionnement »⁵⁹.

La méthode permet ainsi d'étudier des décalages entre modèles, c'est-à-dire des décalages entre modes de saisie des objets. C'est pourquoi ce type d'étude repose essentiellement sur l'analyse des zones de « tension » :

Lorsqu'un modèle formel et une catégorisation pratique se disputent le même objet, il s'établit entre eux une tension dialectique : chacun des deux tend à réduire l'autre par absorption de certains de ses schémas.

Berrendonner et Reichler-Béguelin, 1989, p. 111.

B- Méthode d'analyse du décalage linguistique en indexation

Berrendonner et Reichler-Béguelin adoptent une stratégie d'analyse en termes de décalage pour atteindre ce que les sujets parlants considèrent comme des mots, des phrases, etc., et non pour vérifier la compatibilité de leur vision de linguistes avec celle des alphabètes lambda⁶⁰. Leur approche repose sur l'hypothèse que « les unités non formelles, créées, récupérées ou bricolées à des fins techniques, portent la trace d'un mode de structuration pragmatique du langage obéissant à des règles propres, en vertu de finalités spécifiques »⁶¹. De la même façon, nous pensons que les objets spécifiques à l'indexation (descripteur, langage documentaire par exemple) portent les traces d'une certaine vision de la langue en même temps que celles d'un certain usage de la langue.

C'est pourquoi la méthode d'analyse du décalage nous paraît pouvoir être utilisée :

⁵⁹. Berrendonner et Reichler-Béguelin reprennent ici une distinction proposée par Amalberti [1987].

⁶⁰. Berrendonner et Reichler-Béguelin 1989, p. 103.

⁶¹. *Ibid*, p. 100.

d'une part, pour dégager les représentations linguistiques sous-jacentes à la pratique d'indexation ; d'autre part, pour montrer comment ces représentations peuvent, pour certaines d'entre elles, se constituer en fondements théoriques. Le passage du niveau des représentations à celui des fondements n'est en effet pas automatique : seules quelques représentations linguistiques à l'œuvre en indexation relèvent d'un modèle de fonctionnement formel de la langue et peuvent à ce titre prétendre au statut de fondements théoriques (*i.e.* fondements du point de vue de la théorie linguistique). Dans d'autres cas, les représentations linguistiques sous-jacentes en indexation se révèlent être des représentations *ad hoc* établies *a posteriori* par la pratique, qui n'ont plus alors qu'un lointain rapport avec un quelconque modèle de la langue.

En effet, si les représentations non formelles de la langue en indexation se trouvent régulièrement en conflit avec les représentations formelles des linguistes, il arrive aussi qu'elles les « absorbent » complètement. Dans les termes de Berrendonner et Reichler-Béguelin, il est en effet inévitable que, se disputant les « mêmes objets », les deux types de modèle cherchent à s'absorber et à se confondre ; un modèle d'utilisation de la langue finit par se donner pour un modèle de fonctionnement, alors que le modèle d'utilisation devrait uniquement « utiliser », c'est-à-dire, étendre, déformer, etc., les éléments d'un modèle formel :

La récurrence obstinée de ces polyvalences où sont mis en jeu deux ou plusieurs niveaux d'articulation du langage théoriquement distincts, est à mettre au compte d'une logique utilitaire construisant et adaptant ses unités en fonction de pertinences opératoires multiples. Ces chevauchements sont intrinsèquement liés à l'élaboration spontanée des outillages graphiques et **par contrecoup, ils conditionnent forcément la perception non formelle que les sujets ont des unités de leur langue.**

Berrendonner et Reichler-Béguelin, 1989, p. 102 (c'est nous qui soulignons).

Autrement dit, le modèle d'utilisation de la langue outre qu'il « pervertit »⁶² régulièrement, pour ses besoins propres, les représentations formelles, se place en constante concurrence avec le modèle de fonctionnement de la langue. C'est par cette double tension que nous pourrions montrer comment se constituent les représentations linguistiques en indexation et comment seulement certaines d'entre elles peuvent fonctionner comme des fondements théoriques.

⁶². *Ibid.*, p. 111 : « Les catégorisations pratiques incorporent chroniquement - en les pervertissant - des notions scientifiques vulgarisées ».

Par le biais de la méthode d'analyse du décalage, nous chercherons donc à :

- (i) dégager, à partir de la *doxa*⁶³ linguistique en indexation, le modèle d'utilisation de la langue mis en œuvre, ainsi que le modèle de fonctionnement de la langue auquel il se réfère implicitement : à cette étape, on cherche à dégager les représentations de la langue qui sous-tendent la pratique de l'indexation ;
- (ii) confronter le modèle de fonctionnement implicite en indexation au modèle de fonctionnement explicite en linguistique : à cette étape, on cherche à évaluer la validité des représentations de la langue véhiculées en indexation au regard des hypothèses linguistiques. Les représentations de la langue en indexation sont ou pas conformes aux représentations formelles des linguistes : si elles le sont, elles sont considérées comme constituant des fondements théoriques ; si elles ne le sont pas, on procède à un déplacement de modèle de fonctionnement ;
- (iii) substituer, aux représentations linguistiques de l'indexation non valides sur le plan linguistique, des représentations linguistiques valides, qui constituent alors des fondements théoriques pour l'indexation ; dans le cadre de ces nouvelles représentations, on reformule le modèle d'utilisation de la langue en indexation. À cette étape, on propose une redéfinition de l'indexation (ou plutôt de certains aspects de l'indexation) où les références et les emprunts à la linguistique sont conformes à un modèle théorique de la langue.

On peut schématiser la méthode que nous adopterons pour constituer les fondements de l'indexation du point de vue de la théorie linguistique de la façon suivante⁶⁴ :

⁶³. On rappelle que la notion de « *doxa* » a été proposée par Aristote dans le cadre de ce qu'on a pu appeler une théorie des lieux du langage. La *doxa* constitue le second de ces lieux, c'est le lieu « commun » à tous les sujets parlants qui, à ce titre, peut être vu comme un « tissu de conjectures, d'usages habituels, de comportements les plus ordinaires, de discours vraisemblables », Cauquelin 1990, p. 66-67. Sa principale caractéristique est de constituer la « matière première de l'entente, voire de la concorde, parce que sans elle il n'y aurait rien à partager, rien non plus à préciser, à distribuer selon les lieux », *ibid*, p. 34. C'est pourquoi la *doxa* n'intéresse pas directement par la validité des propos qui s'y tiennent.

⁶⁴. Il nous semble que la méthode adoptée ici peut être aussi utilisée pour la constitution des fondements de l'indexation du point de vue d'autres théories (cognitive, sociologique, etc.).

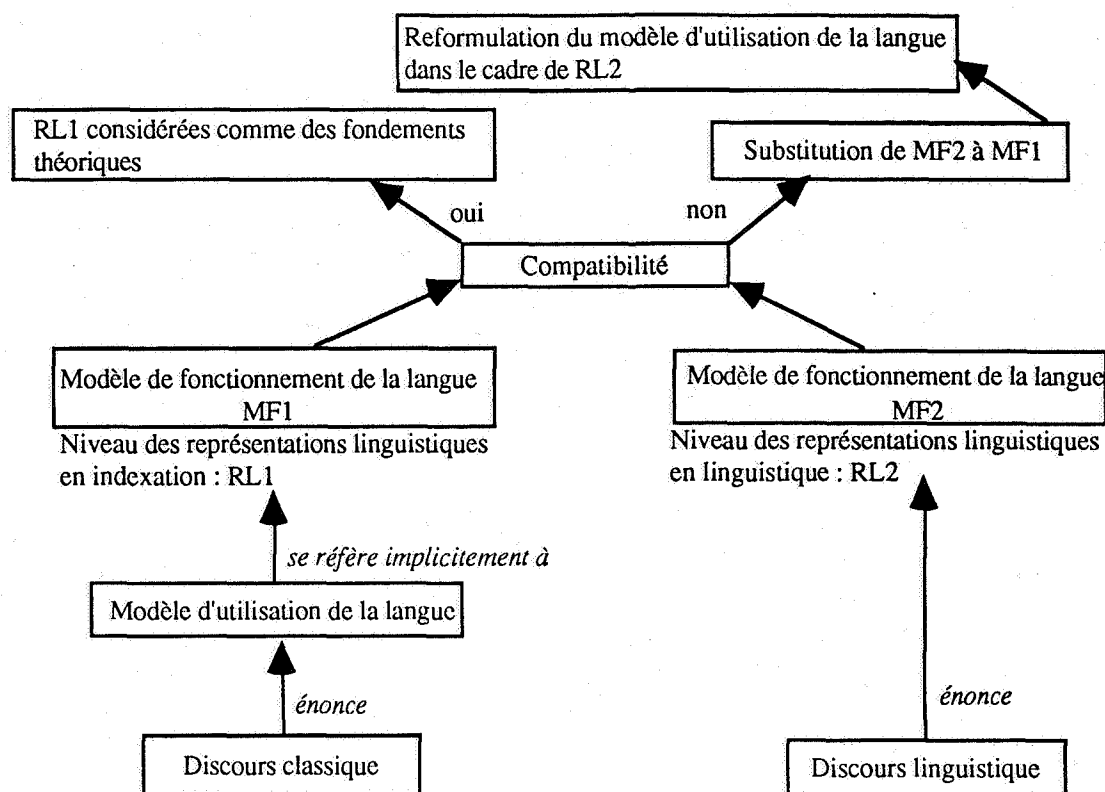


Figure 2- La méthode d'analyse du décalage

Il importe de noter que l'adoption d'une telle méthode favorise, d'une certaine façon, le « détournement » des concepts scientifiques. Sur ce point, il est clair qu'une étude des fondements théoriques de l'indexation telle que nous la proposons (par opposition à une étude qui chercherait à constituer une théorie de l'indexation par exemple) ne permet pas de sortir à proprement parler du « bricolage théorique » qui caractérise la pratique d'indexation. L'objectif consiste simplement à se doter d'un référentiel (ici une théorie linguistique) qui, en étant étranger au domaine documentaire lui-même, permette de sortir du discours de la pratique sur elle-même, et, de ce fait, d'objectiver certains faits d'indexation. Le référentiel proprement documentaire, le discours classique, en ne s'inscrivant dans le cadre d'aucune théorie, ne peut permettre de dépasser la complexité d'une pratique et de « désintriquer » les différents phénomènes en jeu. Par le référentiel linguistique, on pourra faire apparaître une partie des mécanismes de l'indexation : ceux qui reposent sur les propriétés de la langue.

Pour rendre complète la présentation de la méthode d'analyse que nous suivrons dans cette recherche, nous devons déterminer les objets de l'indexation qui peuvent être étudiés dans le cadre proposé. Nous en avons dégagé cinq, susceptibles de nous permettre d'énoncer les fondements de l'indexation du point de vue d'une

théorie linguistique. Les objets de l'indexation ne correspondant pas aux objets de la linguistique, nous présentons dans le tableau suivant, au regard des objets documentaires retenus pour étude, les objets formellement constitués par les linguistiques, qui nous permettront d'avancer dans notre recherche :

Objets de l'indexation	Objets formels
descripteur	- terme textuel - groupe nominal - synapsie
document	- référent - énoncé
information	- objet de discours - thème discursif
langage documentaire	- discours [documentaire]
collection documentaire	- formation discursive

Les objets de l'indexation du tableau ci-dessus ne seront étudiés ni successivement ni séparément, ceci pour deux raisons : d'une part, ils mettent souvent en jeu les mêmes problématiques linguistiques (la référence par exemple pour l'étude du descripteur et du document), d'autre part, ils sont fortement liés entre eux (document et collection documentaire par exemple). Ces différents objets seront plutôt étudiés selon le point de vue qui rend visibles les « lieux de tension » entre les deux modèles formel et non formel de la langue : le lexique, la référence, le discours. Seul le descripteur, parce qu'il permet de rendre compte globalement de l'indexation fera l'objet d'un chapitre particulier, dans lequel nous pourrons synthétiser l'ensemble de nos propositions.

C- Limites et contours de la recherche

Le cadre méthodologique que nous nous proposons de suivre ne permet évidemment pas de traiter tous les aspects de l'indexation : comme nous l'avons vu, on ne peut constituer un modèle d'utilisation qu'à partir d'une seule théorie à la fois. Or, nous avons épousé l'hypothèse que la pratique d'indexation, comme la plupart des pratiques, devait relever de plusieurs champs disciplinaires : à ce titre, elle met en jeu des représentations théoriques de nature différente. Par ailleurs, il n'est pas exclu qu'une même représentation puisse relever de modèles formels différents (par exemple, la notion de « concept » en indexation doit sans doute reposer à la fois sur des représentations linguistiques et sur des représentations cognitives).

Il importe donc d'essayer de formuler les limites qui, à plusieurs niveaux, sont impliquées par le cadre méthodologique retenu :

- limite relative à la couverture de l'objet d'étude : dans l'approche linguistique de l'indexation que nous avons adoptée, nous ne pouvons étudier que des objets documentaires où la matière linguistique se trouve réalisée⁶⁵, c'est-à-dire essentiellement les descripteurs, les documents et les collections documentaires, les langages documentaires et l'information⁶⁶. Il y a donc forcément de nombreux aspects de l'indexation dont cette recherche ne pourra rendre compte⁶⁷ ;
- limite relative à la couverture des objets étudiés : les objets empiriques de l'indexation que nous avons retenus pour notre étude mettent en œuvre, de façon évidente pour certains d'entre eux (comme la notion d'information), plusieurs dimensions et relèvent à ce titre de modèles formels différents. Il est, en l'état actuel, délicat de préciser la couverture descriptive qu'offre, pour un objet donné, une approche linguistique. En conclusion de cette recherche, nous essayerons d'indiquer, au moins pour les objets documentaires que nous aurons étudiés, les autres représentations théoriques qui nous semblent pouvoir être en jeu en indexation.

Les contours de cette recherche restent délicats à déterminer : on distingue mal ce que l'approche linguistique est à même de traiter et ce dont elle ne peut rendre compte. Nous pensons qu'une telle difficulté, liée en grande partie à notre approche de l'indexation, est par ailleurs amplifiée par la « nature » même du langage : s'il est clair que l'on peut, en plus d'un point d'une pratique, relever « des interventions en langue naturelle »⁶⁸, il n'est pas toujours évident de décider celles qui relèvent, « en droit », d'une approche linguistique et celles qui n'en relèvent pas. Sans doute aurons-nous laissé de côté, dans cette recherche, des aspects de l'indexation qu'une linguistique pourrait permettre de traiter, tandis que nous aurons parallèlement préjugé des fondements linguistiques de tels autres aspects de l'indexation. Là encore, il nous semble que la méthode d'analyse proposée permet d'introduire des modifications (par exemple dans les objets d'étude) sans que la

⁶⁵. Des « réalisations de langue » dans la terminologie de Milner 1989, p. 35.

⁶⁶. Pour ce dernier objet, plus que pour les autres, la réalisation de la matière linguistique n'est pas évidente et l'hypothèse de la représentation linguistique que nous ferons (sous la notion d'« objet de discours ») sera alors ici plus forte qu'ailleurs. Nous espérons pouvoir montrer que cette hypothèse n'est cependant pas dénuée d'intérêt.

⁶⁷. Par exemple, tous les aspects qui concernent la façon dont les indexeurs indexent ; sur ce point, on peut se reporter à l'analyse menée dans le cadre des sciences cognitives par Bertrand 1993.

⁶⁸. Milner 1989, p. 32.

démarche elle-même ne soit remise en cause.

Cependant - et ce sera là la dernière précision que nous apporterons à la méthode d'analyse adoptée -, ce n'est pas parce « l'intervention en langue naturelle », pour reprendre les termes de Milner, n'est pas immédiate que le modèle linguistique n'est plus approprié à la description des faits d'indexation. Sur ce point, on tiendra que notre approche linguistique permet de traiter l'indexation de tous les types de documents, que ceux-ci soient de nature textuelle ou non⁶⁹.

En effet, comme cette recherche se donne pour objectif de le montrer, la pratique de l'indexation met toujours en œuvre des discours, quel que soit le type d'objet manipulé (texte, image, son, etc.). Ces discours sont soit directement liés au tissu textuel du document à analyser : ils sont alors particulièrement peu visibles. Ces discours sont, dans le cas des objets non textuels, créés spécifiquement par et/ou pour l'indexation : ils offrent alors une meilleure visibilité. En effet, un objet documentaire à indexer est toujours accompagné d'un texte⁷⁰ : la présence d'un texte est une condition pour qu'un document constitue un objet documentaire⁷¹. Si ce texte n'existe pas, il faut le créer⁷². Nous aurons à expliquer cette présence obligatoire du texte dans l'objet à indexer.

Que les documents non textuels soient pris en compte dans notre approche linguistique de l'indexation ne signifie pas, pour autant, que nous ne tenons pas compte de la spécificité de ce type de documents ; tout au contraire. C'est parce que notre cadre d'analyse est un cadre linguistique que nous pouvons poser ces différences. En effet, dans le cadre d'une approche sémiologique⁷³ telle que Benveniste⁷⁴ a pu la définir, texte et image (par exemple) constituent deux modes

⁶⁹. L'ensemble des objets susceptibles d'être indexés est très divers et mal défini. On suppose que l'indexation est en principe exercée au moins sur l'ensemble des productions éditoriales qui font l'objet d'un dépôt légal. La loi n° 92-546 du 20 juin 1992 relative au dépôt légal français circonscrit le type de productions concernées. Sont ainsi candidats à l'indexation : les documents imprimés, graphiques, photographiques, sonores, audiovisuels, multimédias ; les progiciels, les bases de données, les systèmes experts et les autres produits de l'intelligence artificielle.

⁷⁰. Le cas exemplaire est celui du document iconographique.

⁷¹. Voir la norme de catalogage des images fixes (Norme AFNOR Z 44-077, 1995) qui stipule que « l'image est toujours accompagnée d'un texte. [...] Le texte d'accompagnement est le document historiquement attesté par lequel l'image est identifiée ». Voir aussi Moles : « Une image sans légende de la part de son fournisseur ou producteur n'a pas - en principe et en général - droit d'accès à l'iconothèque », Moles cité in Le Guern Odile 1989, p. 427.

⁷². Voir par exemple Escarpit 1991, p. 161 : « On notera que le documentaliste part toujours d'un texte. Si ce texte n'existe pas, il faut d'abord le constituer. Nous avons vu par exemple que l'analyse directe de l'image pose des problèmes insurmontables. Toute analyse documentaire d'une image doit donc commencer par la production d'un discours descriptif de l'image et la notation de ce discours dans un texte qui constitue le document relais à analyser ».

⁷³. La sémiologie (ou sémiotique) étudie précisément les liens entre systèmes sémiotiques, en se donnant pour objet « l'étude des signes et des processus interprétatifs », Ducrot et Schaeffer 1995, p. 179.

⁷⁴. Benveniste 1974 [1969], p. 43-66.

sémiotiques différents, non réductibles l'un à l'autre⁷⁵. Le texte n'est donc pas, avec l'image, en relation de commentaire ou d'interprétation.

Nous aurons donc à déterminer le type de relation que le texte établit avec l'image en indexation mais aussi l'utilité, pour l'indexation, de disposer d'un texte qui ne renvoie pas au « contenu » de l'image. Cet aspect nous intéresse au plus au point puisqu'il met au jour un « paradoxe » : l'indexation porte, dans ce cas, non sur l'image elle-même mais sur le texte qui l'accompagne (sans l'interpréter), alors même que l'indexation est supposée réaliser une « analyse du contenu » du document (ici un document iconographique). On comprend donc que, dans le cadre de notre analyse des notions de « document » et de « contenu », l'étude des documents non textuels, dans leur non-spécificité textuelle, doit être prise en compte.

Reste cependant que, essentiellement pour des commodités d'exposition, nous nous appuierons, en grande partie, sur des exemples d'indexation portant sur des documents textuels.

En outre ne seront pris en compte dans cette recherche que les textes de documents écrits en français ; cette restriction est liée aux types de théories linguistiques sollicitées qui travaillent essentiellement les phénomènes linguistiques du français. La dimension multilingue, nécessaire pour aborder l'indexation en toute généralité, pose des problèmes spécifiques qu'il ne nous est pas possible d'appréhender dans cette recherche.

III- Synthèse du chapitre et présentation du plan de l'étude

III.1- Synthèse du chapitre

Dans ce chapitre, nous avons défini la problématique de notre recherche en la situant à deux niveaux : celui de l'objet d'étude et celui de la méthode d'analyse.

À chaque niveau, nous avons pu relever, dans les approches classiques de l'indexation, des impasses ou des contradictions, à partir desquelles nous avons formulé nos propres hypothèses. On rappelle ci-dessous l'essentiel de l'argumentation présentée :

- une approche de l'indexation définie en termes de recherche documentaire ne permet pas de conduire une étude de l'indexation sur un plan théorique, ceci pour deux raisons : d'une part, l'objet

⁷⁵. « Il n'y a pas de "synonymie" entre systèmes sémiotiques ; on ne peut pas "dire la même chose" par la parole et par la musique, qui sont des systèmes à base différente ». *Ibid.* p. 53.

indexation n'y est pas autonomisable ; d'autre part, les présupposés des approches classiques de l'indexation conduisent à une impasse et à une contradiction : supposée être une technique, l'indexation devrait pouvoir faire l'objet d'une théorie, or c'est une théorie de la recherche documentaire qui fait l'objet de recherche ;

- en posant une distinction entre indexation et recherche documentaire, on fait l'hypothèse que l'indexation peut être appréhendée de façon autonome. En posant une distinction entre technique et pratique, on fait l'hypothèse que l'indexation, entendue comme une pratique, si elle ne peut faire l'objet d'une théorie, repose sur des représentations théoriques, qui peuvent se constituer en fondements théoriques, c'est-à-dire en fondements du point de vue d'une théorie. Là encore, deux hypothèses : d'une part, que l'indexation est une pratique empruntant à différentes théories ses modèles de fonctionnement ; d'autre part, que, parmi les représentations théoriques sous-jacentes à la pratique de l'indexation, les représentations linguistiques sont pertinentes pour constituer les fondements théoriques de l'indexation.

Dans le cadre de notre recherche, l'indexation, entendue comme une pratique reposant sur des fondements théoriques, est donc considérée de la façon suivante :

- l'indexation porte uniquement sur des documents mais sur tous les types de document, quelle que soit leur nature, textuelle ou non ; en revanche, on ne prend pas en compte dans cette recherche la notion d'indexation de requêtes ;
- l'indexation est à entendre comme un processus au sens « fort » : elle réalise des opérations de fabrication de ses objets, notamment les documents et les informations ; cette recherche devra montrer que la notion de processus au sens « faible » (notion de transfert) est insuffisante ;
- l'indexation se réalise par d'autres moyens que les « mots » : la notion de langage documentaire est à mettre à distance au profit d'autres procédés non plus lexicaux mais discursifs. Cette recherche devra montrer le caractère réducteur d'une approche de l'indexation qui ne considérerait que les « mots ».

Le cadre d'analyse retenu permet de donner corps à une telle approche de l'indexation ; en effet, par déplacements successifs de modèles, il donne la possibilité de :

- montrer comment, sur des points de conflits précis entre modèles formel et non formel de la langue, la description classique de l'indexation achoppe ;
- reconstruire, sur la base de ces lieux d'achoppement, de nouvelles représentations qui permettent de faire « voir » les fondements théoriques de l'indexation.

Cette étude est donc construite autour de deux principaux axes que nous présentons ci-après.

III.2- Présentation du plan de la recherche

La recherche est menée sur quatre chapitres, regroupés en deux parties distinctes : une première partie essentiellement critique ; une seconde partie plus prospective. Chaque partie faisant l'objet d'une introduction spécifique, on en présente ici simplement les grandes lignes.

Première partie : les problèmes théoriques de l'indexation

Cette première partie emprunte son titre et son esprit à un ouvrage de Georges Mounin paru en 1963⁷⁶, dont la problématique nous paraît être très proche de la nôtre. Mounin y montre en effet en quoi l'activité professionnelle de la traduction constitue un « scandale » au regard des théories linguistiques⁷⁷ et pose les différentes pistes de recherche possibles une fois ce constat posé⁷⁸ :

- (i) on peut remettre en cause le bien fondé d'une pratique ;
- (ii) on peut remettre en cause le bien fondé des théories linguistiques ;
- (iii) on peut chercher à comprendre les raisons d'une cohabitation entre impossibilité théorique et possibilité empirique.

Si, tout comme Mounin, nous considérons que seule la voie (iii) est constructive, il importe de commencer par mettre au jour les incompatibilités entre modèle d'utilisation de la langue en indexation et modèle de fonctionnement de la langue en

⁷⁶. Mounin 1963 : Les Problèmes théoriques de la traduction.

⁷⁷. Mounin 1963, p. 8 : « On pourrait presque dire que l'existence de la traduction constitue le scandale de la linguistique contemporaine ».

⁷⁸. *Ibid.*, p. 8-9.

linguistique.

Pour mettre au jour ces incompatibilités, on s'intéresse aux lieux de « tension » entre ces deux modèles (les lieux où les modèles se « disputent les mêmes objets » dans les termes de Berrendonner et Reichler-Béguelin) ; nous en avons identifié deux, que nous traitons respectivement dans les chapitres II et III qui constituent cette première partie : le lexique (chapitre II) et la référence (chapitre III).

Deuxième partie : contribution aux fondements théoriques de l'indexation

La première partie fait apparaître que le modèle de fonctionnement de la langue sous-jacent à la pratique de l'indexation s'apparente, sur les questions du lexique et de la référence, au modèle du sens commun, contre lequel s'est notamment construite la linguistique. On peut choisir d'en rester là, mais on ne peut alors dégager des fondements théoriques de l'indexation.

On peut aussi envisager, dans la perspective de constituer les fondements de l'indexation du point de vue de la théorie linguistique, de substituer à ce modèle de sens commun un modèle scientifique : c'est ainsi que l'on propose, dans cette seconde partie, un nouveau modèle d'utilisation de la langue qui s'inscrit, lui, dans le cadre d'un modèle linguistique. On est amené à redéfinir l'indexation sous ses deux principaux aspects, processus et résultat :

- appréhendée sous l'angle du processus, l'indexation est analysée dans une perspective discursive : cette étude fait l'objet du chapitre IV ;
- appréhendée sous l'angle du résultat, l'indexation est réinterrogée dans ses « formes » : formes logique et linguistique du descripteur, traitées dans le chapitre V.

Cette deuxième partie propose des éléments pour constituer les fondements théoriques de l'indexation, éléments qui, établis dans le cadre d'un modèle d'utilisation de la langue, relèvent d'emprunts et de transformations de concepts issus de la linguistique. Sur ce point, le modèle d'utilisation de la langue proposé en deuxième partie, s'il ne sort pas à proprement parler du « bricolage », permet de souligner ce que l'indexation gagne à exploiter l'« hétérogénéité » des textes et les « ambiguïtés » de la langue.

Nous espérons ainsi montrer en quoi le modèle d'analyse adopté peut se révéler fécond. Il permet de penser plusieurs types d'approche de l'indexation : l'indexation classique qui se greffe, faute de mieux et par commodité, sur le modèle

du sens commun, mais qui du même coup se heurte, sans pouvoir les contourner, aux propriétés de la langue ; une approche discursive de l'indexation qui cherche à prendre en compte, à tirer partie de la langue, ambiguë, variée, hétérogène.